

## CARDIOLOGIA BASEADA NA EVIDÊNCIA

# Análise de Subgrupos em Ensaios Clínicos Terapêuticos [26]

ANTÓNIO VAZ CARNEIRO

Centro de Estudos de Medicina Baseada na Evidência da Faculdade de Medicina de Lisboa, Lisboa

Rev Port Cardiol 2002;21 (3):339-346**RESUMO**

A prática da terapêutica cardiológica deve ser baseada em evidência científica sólida, obtida preferencialmente em ensaios clínicos aleatorizados, prospectivos e controlados (*randomized controlled trials* – RCTs), já que é este o desenho do estudo que permite um estabelecimento mais rigoroso da causalidade. A aplicação dos resultados de

RCTs ao doente individual implica um conjunto de regras específicas, mas que – basicamente – podem ser resumidas na questão «...será que o meu doente poderia ter entrado neste estudo?...». Se a resposta a esta questão da elegibilidade do doente for positiva, então a possibilidade de aplicação dos resultados do ensaio é maior. Se for negativa, deverá o cardiologista agir com mais prudência na sua decisão terapêutica.

Num RCT – seja de que dimensão for – é quase sempre possível identificar sub-grupos de doentes que, fazendo parte da amostra global, sugerem podermos estar perante diferentes efeitos do tratamento: por exemplo, estudos provam que, na profilaxia do acidente vascular cerebral em doentes com fibrilhação auricular de origem não-reumática, a anticoagulação é mandatória em todos os doentes, excepto nos que têm menos de 65 anos e nenhum risco cardiovascular adicional, caso em que bastará antiagregar o doente com aspirina.

Quando se detecta uma diferença no efeito do tratamento entre sub-grupos que é simultaneamente real e de dimensão apreciável, então este achado pode influenciar o modo como se trata o doente individual – daqui a importância da análise de sub-grupos nos RCTs. Esta, no entanto, deverá ser feita com as maiores precauções,

**ABSTRACT****Subgroup Analysis  
in Therapeutic Trials**

Therapy in cardiology must be based on solid scientific evidence, obtained in randomized controlled trials (RCTs), since this is the best design that proves causality in medicine. The applicability of clinical trial results to the individual patient depends on a rigorous set of rules that can be summarized in the question “Could my patient have been enrolled in this trial?” If the answer to this question is affirmative, then the possibility of applying the trial results is greater. If it is negative, then the cardiologist should exercise caution in his or her decision.

In an RCT – of whatever size – it is almost always possible to identify subgroups of patients that show significant differences in treatment effect: for example, studies have shown that, in patients with non-rheumatic atrial fibrillation, oral anticoagulants should be given to prevent stroke, except in those younger than 65 years with no additional risk factors, for whom aspirin is a better option. Subgroup analysis is important because, when the magnitude of the difference is both real and large, it may influence patient management. This analysis should be done with great care, since it has the potential to lead to major errors in data interpretation, identifying differences in treatment effects that are due to chance alone or, more frequently, have no clinical significance. In this article we present a set of guidelines that enable the cardiologist to assess the credibility of an analysis that shows apparent differences in treatment effects across subgroups.

já que pode induzir em erros graves, identificando diferenças terapêuticas resultantes de mero acaso ou, como mais frequentemente acontece, clinicamente irrelevantes.

No presente artigo iremos apresentar um conjunto de normas que permitem ao cardiologista identificar com segurança quais os sub-grupos credíveis em termos de resultados terapêuticos.

#### Palavras-Chave

Análise de sub-grupos;  
Ensaio aleatorizado prospectivo e controlado; Viéses;  
Cardiologia baseada na evidência

#### Key words

Sub-group analysis; Randomized controlled trials;  
Bias; Evidence-based cardiology

## INTRODUÇÃO

A prática da terapêutica cardiológica deve ser baseada em evidência científica sólida, obtida preferencialmente em ensaios clínicos aleatorizados, prospectivos e controlados (*randomized controlled trials* – RCTs) de boa qualidade metodológica<sup>(1)</sup>. Com efeito, é este desenho de estudo que permite um estabelecimento mais rigoroso da causalidade, quer numa manobra preventiva (rastreamento de hipercolesterolemia, por exemplo), quer numa terapêutica (utilização de inibidores do enzima de conversão na insuficiência cardíaca por disfunção sistólica do ventrículo esquerdo, por ex.).

A aplicação dos resultados de RCTs ao doente individual implica um conjunto de regras específicas, mas que – basicamente – podem ser resumidas na questão «...será que o meu doente poderia ter entrado neste estudo?...». Se a resposta a esta questão da elegibilidade do doente for positiva, então a possibilidade de aplicação dos resultados do ensaio é maior. Se for negativa, deverá o cardiologista agir com mais prudência na sua decisão terapêutica<sup>(2,3)</sup>.

Num RCT – seja de que dimensão for – é quase sempre possível identificar sub-grupos de doentes que, fazendo parte da amostra global, sugerem podermos estar perante diferentes efeitos do tratamento<sup>(4)</sup>. Deste modo, uma análise de sub-grupos positiva – isto é, sugerindo diferenças terapêuticas – permite um potencial refinamento de aplicação de resultados ao doente individual: por exemplo, na profilaxia do acidente vascular cerebral em doentes com fibrilhação auricular de origem não-reumática, a anticoagulação é mandatória em todos os doentes, excepto nos que têm menos de 65 anos e

## INTRODUCTION

Therapy in cardiology must be based on solid scientific evidence, ideally obtained in prospective randomized controlled trials (RCTs) of sound methodology<sup>(1)</sup>, since this is the best design that proves causality in medicine, whether in preventive action (screening for high blood cholesterol levels, for example), or in treatment (such as the use of angiotensin-converting enzyme inhibitors for heart failure due to left ventricular systolic dysfunction).

The applicability of clinical trial results to the individual patient depends on a rigorous set of rules that can basically be summarized in the question “Could my patient have been enrolled in this trial?” If the answer to this question as to the eligibility of the patient is affirmative, then the possibility of applying the trial results is greater. If it is negative, then the cardiologist should exercise more caution in his or her decision<sup>(2,3)</sup>.

In an RCT – of whatever size – it is almost always possible to identify subgroups of patients within the overall sample that show significant differences in treatment effect<sup>(4)</sup>. Thus, a positive subgroup analysis – i.e. one suggesting treatment differences – enables the application of trial results to be tailored to the individual patient. For example, studies have shown that oral anticoagulants should be given to all patients with non-rheumatic atrial fibrillation to prevent stroke, except patients younger than 65 years with no additional risk factors, for whom aspirin is sufficient<sup>(5)</sup>.

Differences in treatment effects in subgroups of an RCT can be found in three contexts:

nenhum risco cardiovascular adicional, caso em que bastará antiagregar o doente com aspirina<sup>(5)</sup>.

As diferenças nos efeitos terapêuticos encontradas em sub-grupos de um RCT podem verificar-se em três contextos:

- nos doentes propriamente ditos (por ex. homens ou mulheres, grupos etários diversos, diferentes raças);
- nos esquemas terapêuticos utilizados (diversas vias de administração ou dosagens do mesmo fármaco, diferentes medicamentos da mesma classe, etc.);
- na medição dos resultados (*outcomes*) do ensaio (mais precoces ou mais tardios, definição de limiares de significação clínica, etc.).

Por exemplo: a utilização de terapêutica anti-dislipidémica reduz para metade o risco de doença coronária aos 10 anos num doente hipertenso, diabético, fumador e idoso (que apresenta um risco de base de 40 %), assim como numa doente jovem, com único factor de risco a hipercolesterolemia (risco de base de 4 %). No primeiro doente, a redução do risco absoluto pela terapêutica é de 20 %, enquanto que na mulher jovem será, quando muito, de 2 %. Este exemplo demonstra a relação entre um estrato de risco cardiovascular e um efeito terapêutico.

Resumindo: quando se detecta uma diferença no efeito do tratamento entre sub-grupos que é simultaneamente real e de dimensão apreciável, então este achado pode influenciar o modo como se trata o doente individual - daqui a importância da análise de sub-grupos nos RCTs. Esta deverá ser feita com as maiores precauções, já que pode induzir erros graves de interpretação dos dados, identificando diferenças terapêuticas resultantes de mero acaso ou, como mais frequentemente acontece, clinicamente irrelevantes<sup>(6)</sup>.

No presente artigo iremos apresentar um conjunto de normas que nos permitem identificar com segurança quais os sub-grupos credíveis em termos de resultados terapêuticos, reconhecendo que este método pode revelar-se falível em ensaios com amostras pequenas e poder analítico baixo (existência de número apreciável de falsos-negativos).

## REGRAS METODOLÓGICAS DE ANÁLISE

O quadro seguinte descreve o conjunto de questões metodológicas que devem ser respondidas afirmativamente para que a análise de subgrupos de um RCT seja credível.

- in the patients themselves (e.g. men *vs* women, different age groups, different races);
- in the treatment schemes used (different dosages or methods of administering the same drug, different drugs of the same class, etc.);
- in the evaluation of the outcomes of the trial (earlier or later, definition of cutoff points of clinical significance, etc.).

As an example, the use of antidyslipidemic therapy halves the risk of coronary artery disease at 10 years in a hypertensive, diabetic, elderly male smoker (who presents a baseline risk of 40 %) as well as in a young woman whose only risk factor is high blood cholesterol (baseline risk of 4 %). In the first patient the reduction in absolute risk from therapy is 20 %, while in the young woman it is at most 2 %. This example shows the relationship between a cardiovascular risk stratum and a treatment effect.

To summarize, subgroup analysis of RCTs is important because, when the magnitude of a difference detected in treatment effect is both real and large, it may influence individual patient management. The analysis should be done with great care, since it has the potential to lead to major errors in data interpretation, identifying differences in treatment effects that are due to chance alone or, more frequently, that have no clinical significance<sup>(6)</sup>.

In this article we present a set of guidelines that enable cardiologists to assess the credibility of an analysis that shows apparent differences in treatment effects across subgroups, while acknowledging that it may be unreliable in trials with small samples or low analytical power (a significant number of false negatives).

## METHODOLOGICAL GUIDELINES FOR ANALYSIS

The table below sets out the series of methodological questions that need to be answered affirmatively for any subgroup analysis of an RCT to be credible.

**Is the treatment difference found between the subgroups based on comparisons within a single study rather than between several studies?**

The differences found between subgroups within a single study are more credible than

**Questões metodológicas na definição de diferenças reais do efeito terapêutico entre sub-grupos de um ensaio clínico**

- a diferença terapêutica verificada entre os sub-grupos é baseada em comparações encontradas num estudo único, ou entre vários estudos?
- a hipótese da diferença terapêutica entre sub-grupos foi estabelecida previamente à análise dos dados, ou foi estabelecida posteriormente?
- o efeito do tratamento encontrado num subgrupo fez parte de um número relativamente pequeno de hipóteses colocadas *a priori*?
- a magnitude do efeito terapêutico apresenta dimensão apreciável?
- o efeito do tratamento foi estatisticamente significativo?
- o efeito terapêutico foi consistente entre estudos diferentes?
- existe evidência fisiológica indirecta que suporte a diferença de efeito terapêutico?

**A diferença terapêutica verificada entre os sub-grupos é baseada em comparações encontradas num estudo único, ou entre vários estudos?**

As diferenças entre sub-grupos encontradas num único estudo são mais credíveis do que as encontradas em estudos diferentes, devido ao facto de que, neste último caso, as populações serem diferentes em termos de riscos de efeitos secundários, patologias associadas, intervenções concomitantes e medicações associadas, complicando deste modo a análise de sub-grupos<sup>(7)</sup>.

Por exemplo, no tratamento da insuficiência cardíaca congestiva por disfunção sistólica do ventrículo esquerdo, não é aconselhável comparar a dimensão do efeito terapêutico do carvedilol com o bisoprolol, baseado nos dois estudos publicados sobre estas intervenções<sup>(8, 9)</sup>. Para isso, o ideal seria dispor dos dados de um estudo com três braços terapêuticos na mesma amostra: um a que se administrava o carvedilol, outro o bisoprolol e um terceiro o placebo. Quaisquer diferenças assim detectadas poderiam ser interpretadas como sendo devidas apenas ou às propriedades dos fármacos em questão (não pondo de parte o acaso).

**A hipótese da diferença terapêutica entre sub-grupos foi estabelecida previamente à análise dos dados, ou foi estabelecida posteriormente?**

A tecnologia informática permite analisar os dados de um estudo com uma profundidade e um detalhe impossíveis de conseguir antes

**Methodological questions to define real differences in treatment effect between subgroups of a clinical trial**

- Is the treatment difference found between the subgroups based on comparisons within a single study rather than between several studies?
- Was the hypothesis of treatment difference between subgroups formulated before rather than after the data were analyzed?
- Is the treatment effect found in a subgroup one of a relatively small number of *a priori* hypotheses?
- Is the magnitude of the treatment effect large?
- Was the treatment effect statistically significant?
- Was the treatment effect consistent between different studies?
- Is there indirect physiological evidence to support the difference in treatment effect?

those found in different studies, owing to the fact that in the latter case the populations will be different in terms of risk of side-effects, associated pathologies, concomitant therapies and associated medications, which will complicate subgroup analysis<sup>(7)</sup>.

For instance, in the treatment of congestive heart failure due to left ventricular systolic dysfunction, it is inadvisable to compare the extent of the treatment effects of carvedilol with that of bisoprolol on the basis of the two published studies on these therapies<sup>(8, 9)</sup>. In this case, the ideal would be to have data from a three-arm trial on the same sample: one treated with carvedilol, one with bisoprolol, and the third with a placebo. Any differences detected could then be interpreted as due solely to the properties of the drugs under study (without discounting chance).

**Was the hypothesis of treatment difference between subgroups formulated before rather than after the data were analyzed?**

Advances in information processing enable the data from a study to be analyzed in ways, and at a level of detail, that were impossible before the advent of computers. This ability means that relations between data may be discovered that are completely spurious and, in most cases, due to chance. This so-called “fishing for significance” in trial data is, of course, never part of the initial design and planning of the study. Such post-hoc analyses should be treated with suspicion due to their inherent

do advento dos computadores. Este facto permite que, potencialmente, se descubram relações entre dados que são de facto perfeitamente espúrias e, a maior parte das vezes, devidas ao acaso. Esta técnica designa-se por «pesca de significação» nos dados do ensaio e, como é óbvio, nunca é contemplada previamente, quando do desenho e projecção iniciais do estudo: esta análise post-hoc deve ser questionada, já que a sua fiabilidade é escassa. Por exemplo, só recentemente foi corrigida<sup>(10)</sup> uma falácia sobre a ausência de eficácia da aspirina na prevenção secundária do AVC em mulheres, um dado que tinha sido «descoberto» numa associação ocasional num primeiro RCT, através de uma análise post-hoc dos dados<sup>(11)</sup>.

Claro que uma análise post-hoc nem sempre é desprovida de valor: existem casos de ensaios sobre cujos dados foram aplicados diversos esquemas de análise *a posteriori* e que detectaram relações inesperadas, que serviram de hipótese de pesquisa para novos ensaios prospectivos, cujos resultados foram concordantes com a análise referida, validando-a integralmente. Mas, como regra, devemos suspeitar da real existência de significação de relações feitas depois do estudo ter acabado, e não contempladas antes do seu início<sup>(12)</sup>.

### **O efeito do tratamento encontrado num subgrupo fez parte de um número relativamente pequeno de hipóteses colocadas a priori?**

Quanto maior for o número de hipóteses testadas num RCT, maior o número de interações descobertas pela acção do puro acaso – quer os investigadores as tenham definido *ab initio*, quer não. Infelizmente, na maior parte das vezes, o número de hipóteses que foi realmente testada não é descrita no estudo, pelo que se torna difícil esta análise.

Dois exemplos: o primeiro é o de um estudo em que os investigadores fizeram uma análise de regressão sobre 16 variáveis preditivas do efeito terapêutico da digoxina em doentes com insuficiência cardíaca congestiva e em ritmo sinusal<sup>(13)</sup>. Entre outras, os investigadores chegaram à conclusão que um terceiro tom cardíaco era bom preditor da eficiência da digoxina neste contexto. Dado que este achado nunca foi ulteriormente confirmado em outros estudos sobre a mesma temática, parece razoável duvidar deste achado (um entre dezasseis).

O segundo exemplo é o da análise de 146 subgrupos feita pelos investigadores do *Beta-Blocker Heart Attack Trial* – BHAT –<sup>(14)</sup> que

unreliability. An example of this is the recent correction<sup>(10)</sup> of a fallacy concerning the inefficacy of aspirin for secondary stroke prevention in women, a finding that had been “discovered” in a chance association in a first RCT through post-hoc analysis of the data<sup>(11)</sup>.

Of course, post-hoc analyses are not always without value. There are examples of trials in which the data have been submitted to various types of post-hoc analysis that detected unexpected relationships, which were then used to construct research hypotheses for further prospective trials that produced results in agreement with the initial analysis and completely validating it. However, as a rule, we should be wary of assuming real significance in relationships detected after a study has ended and not planned before it began<sup>(12)</sup>.

### **Is the treatment effect found in a subgroup one of a relatively small number of a priori hypotheses?**

The greater the number of hypotheses tested in an RCT, the greater the number of interactions likely to be discovered due to pure chance – whether or not the investigators defined them *ab initio*. Unfortunately, in most cases, the number of hypotheses that were in fact tested is not stated in the study, which makes it difficult to analyze this particular aspect.

Two examples will illustrate this. The first is a study in which the investigators performed a regression analysis on 16 predictive variables of the therapeutic effect of digoxin in patients with congestive heart failure and in sinus rhythm<sup>(13)</sup>. Among the conclusions arrived at was that a third heart sound was a good predictor of the efficacy of digoxin in this setting. As this has never been confirmed in subsequent studies on the same subject, it seems reasonable to remain doubtful of this finding (one of sixteen).

The second example is of the analysis of 146 subgroups by the investigators of the Beta-Blocker Heart Attack Trial (BHAT)<sup>(14)</sup>, who concluded that, although propranolol was effective overall, there were several subgroups with highly significant variations in treatment effect, in some cases tending towards greater effectiveness, in others the opposite (these results were due to poorly selected samples).

### **Is the magnitude of the treatment effect large?**

The magnitude of the treatment effect in an RCT is defined as the difference in outcomes



concluíram que, embora o propranolol fosse globalmente eficaz, havia vários subgrupos em que o seu efeito terapêutico variava muito significativamente, quer no sentido de uma maior efectividade, quer no contrário (estes resultados eram devido a amostras mal seleccionadas).

### **A magnitude do efeito terapêutico apresenta dimensão apreciável?**

A dimensão do efeito terapêutico num RCT define-se como a diferença dos resultados (*outcomes*) entre o grupo experimental e o de controlo dividido por uma medida de variabilidade (tipicamente o desvio padrão).

Quanto maior a dimensão do efeito terapêutico entre um subgrupo e o resultado global, maior a credibilidade sobre a verdade da sua existência real e o seu significado clínico. No entanto isto não é sempre assim: no caso em que a amostra apresenta dimensão reduzida, pode detectar-se uma diferença marcada entre o efeito aparente que é devida puramente ao acaso<sup>(4)</sup>.

### **O efeito do tratamento foi estatisticamente significativo?**

O estudo GUSTO demonstrou, em doentes com enfarte de miocárdio tratados com tPA versus estreptoquinase, uma redução global de 1 % na mortalidade. A diferença na mortalidade nos EUA foi de 1,2 % e nos outros locais do estudo de 0,7 %<sup>(15)</sup>. Dado que os doentes americanos sofreram durante o estudo maiores taxas de revascularização (angioplastia ou cirurgia de *bypass*), poder-se-ia postular que os benefícios da tPA seriam maiores no contexto de abordagens mais agressivas, isto é, haveria um subgrupo de doentes que beneficiariam mais da trombólise se esta fosse acompanhada de técnicas de revascularização. Como se poderia esclarecer esta dúvida?

A questão mais importante que se deve colocar na análise de diferenças terapêuticas encontradas em sub-grupos de um RCT (partindo do princípio que o efeito é real e comum a todos os doentes) é a da chance que uma diferença entre sub-grupos seja inteiramente devida ao acaso. No caso do estudo GUSTO, a maneira errada seria a de testar primeiro a significação do efeito nos EUA e, independentemente, nos outros países. A maneira correcta envolveria a colocação da seguinte questão analítica: «...na ausência de diferença de efeito terapêutico de base com a tPA *vs* estreptoquinase – a hipótese nula – qual seria a frequência com que se de-

between the experimental group and the control group divided by a measure of variability (normally the standard deviation).

The greater the magnitude of the treatment effect between a subgroup and the overall result, the more likely it is to be real and clinically significant. However, this is not always the case: in a small sample, a marked difference may be detected in apparent effect that is due purely to chance<sup>(4)</sup>.

### **Was the treatment effect statistically significant?**

The GUSTO study, on patients with myocardial infarction treated with tPA versus streptokinase, showed an overall reduction in mortality of 1 %. The difference in mortality in the USA was 1.2 % and in other countries in the study it was 0.7 %<sup>(15)</sup>. Since the American patients in the study underwent higher rates of revascularization (angioplasty or bypass surgery), it could be postulated that the benefits of tPA are greater in a setting of more aggressive intervention; in other words, there may be a subgroup of patients who benefit more from thrombolysis if this is accompanied by revascularization procedures. How can this be confirmed?

The most important question that should be raised in the analysis of treatment differences found in subgroups of an RCT (based on the assumption that the effect is real and common to all the patients) is the likelihood that a difference between subgroups is entirely due to chance. In the case of the GUSTO study, the wrong approach would be to test the significance of the effect in the USA and independently in the other countries. The correct method would involve putting the following analytical question: “In the absence of a difference in baseline treatment effect with tPA *vs* streptokinase – the null hypothesis – how frequently would differences equal to or greater than the estimated 1.2 % and 0.7 % be encountered?”

The answer involves analyzing the level of significance found. Since, in the study, the value of *p* for this test was 0.3, we can state that the data do not support the hypothesis that the effect of tPA was different between the subgroups under consideration (Americans and others). In other words, what the *p* value tells us is that, if the real mortality gradient in the American subgroup and in that of other countries were the same – for example 1 % – there

tectariam diferenças iguais ou superiores às estimadas de 1.2 % e 0.7 %?...»

A resposta passa por analisar o nível de significação encontrado: dado que, no estudo, o valor de P para este teste foi de 0.3, podemos afirmar que os dados não suportam a hipótese que o efeito do tPA seja diferente entre os sub-grupos considerados (EUA e outros). Por outras palavras, o que o valor daquele P nos diz é que, se o verdadeiro gradiente da mortalidade entre os sub-grupos dos EUA e dos outros países fosse idêntico – por ex. 1 % - ocorreriam diferenças iguais ou superiores às observadas em cerca de 30 % do tempo (um resultado que confirma os achados do estudo) <sup>(7)</sup>.

Os leitores interessados em aprofundar esta temática podem consultar fonte que ensina a utilizar uma série de técnicas de análise para detecção do acaso na diferença entre sub-grupos <sup>(16)</sup>.

### **O efeito terapêutico foi consistente entre estudos diferentes?**

Como já foi dito, a credibilidade dos resultados de um efeito terapêutico num subgrupo específico aumenta se este efeito for encontrado em vários estudos diferentes (*versus* se se detectar num estudo único).

De facto, a existência de um efeito de subgrupo em múltiplos estudos constitui a prova mais sólida da sua existência real <sup>(17)</sup>. O estudo mencionado sobre a utilidade do terceiro tom cardíaco como factor preditivo da eficácia da digoxina em doentes em ritmo sinusal é um bom exemplo <sup>(13)</sup>. Este achado foi testado num outro RCT, que não confirmou esta hipótese <sup>(18)</sup>. Assim sendo, o papel do 3º tom cardíaco como preditor de eficácia da digoxina deverá ser encarado com precaução.

### **Existe evidência fisiológica indirecta que suporte a diferença de efeito terapêutico?**

Se os achados de efeitos terapêuticos de sub-grupos forem fisiologicamente plausíveis, será mais fácil aceitá-los como verdadeiros, já que baseados em evidência fisiológica indirecta.

Por «evidência fisiológica indirecta» entendem-se três tipos de provas: 1) existência de interacções em intervenções similares, 2) estudos em populações diferentes (humanas ou animais), e 3) resultados de estudos com outcomes semelhantes (intermediários, por exemplo). Destas, a primeira constitui a prova mais fraca, enquanto que a última é a mais forte.

would be differences that were equal to or greater than those observed about 30% of the time (a result that confirms the findings of the study) <sup>(7)</sup>.

Readers interested in studying this subject further can consult a source that teaches a series of analytical techniques to detect the operation of chance in the differences between subgroups <sup>(16)</sup>.

### **Was the treatment effect consistent between different studies?**

As already stated, the credibility of the results of a treatment effect in a specific subgroup is increased if this effect is found in several different studies (rather than in a single study).

Indeed, if a subgroup effect is found in multiple studies, this is the best evidence that it really exists <sup>(17)</sup>. A good example is the study mentioned above on the usefulness of a third heart sound as a predictive factor for the efficacy of digoxin in patients in sinus rhythm <sup>(13)</sup>. This finding was tested in another RCT, which did not confirm this hypothesis <sup>(18)</sup>. The role of the third heart sound as a predictor of the efficacy of digoxin should thus be treated with caution.

### **Is there indirect physiological evidence to support the difference in treatment effect?**

If the findings of treatment effects in subgroups are physiologically plausible, it is easier to accept them as real, as being based on indirect physiological evidence.

“Indirect physiological evidence” is understood to be of the following three types : 1) interactions in similar interventions; 2) studies in different populations (human or animal); and 3) results of studies with similar (for instance intermediate) outcomes. Of these, the first is the weakest evidence, while the last is the strongest.

The point to remember is that, irrespective of other factors, an interaction between a subgroup of patients and an intervention will always be more credible if it is consistent with current knowledge of the relevant physiological mechanisms <sup>(19)</sup>.

## **CONCLUSIONS**

Subgroup analysis of an RCT is important because, when a difference is detected in the

A mensagem a reter é a de que, independentemente de outros factores, será sempre mais credível uma interacção entre um sub-grupo de doentes e uma intervenção, se esta for consistente com o conhecimento actual dos respectivos mecanismos fisiológicos<sup>(19)</sup>.

## CONCLUSÕES

A análise de sub-grupos de um RCT é importante já que, quando se detecta uma diferença no efeito do tratamento entre sub-grupos este achado pode influenciar o modo como se trata o doente individual. Esta deverá ser feita com precaução, para evitar a aceitação de diferenças terapêuticas resultantes de mero acaso ou, como mais frequentemente acontece, clinicamente irrelevantes.

A utilização do conjunto de normas que foram apresentadas neste estudo podem permitir uma maior certeza na análise de resultados terapêuticos em sub-grupos de doentes num ensaio terapêutico.

treatment effect between subgroups, this finding may influence patient management. The analysis should be done with great care, in order to avoid accepting treatment differences that result from mere chance or, more frequently, have no clinical significance.

Use of the set of guidelines presented in this article may lead to greater certainty in analyzing treatment results in subgroups of patients in a therapeutic trial.

Pedido de separatas para:

Address for reprints:

ANTÓNIO VAZ CARNEIRO

Centro de Estudos de Medicina Baseada na Evidência

Faculdade de Medicina de Lisboa

1649-035 LISBOA

## BIBLIOGRAFIA / REFERENCES

1. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. Evidence-Based Medicine. How to practice and teach EBM. Edinburgh: Churchill Livingstone 2000.
2. Dans AL, Danzl DF, Guyatt GH, Richardson S. The Evidence-Based Medicine Working Group. User's guide to the medical literature. XIV. How to decide on the applicability of clinical trial results to your patients. JAMA 1998;279:545-9.
3. Guyatt GH, Jaeschke RZ, Cook DJ. Applying the findings of clinical trials to individual patients. ACP J Club 1995; Mar-Apr:A12-A13.
4. Detsky AS, Naglie IG. Subgroup analysis: primary and secondary. ACP J Club 1995;May-Jun:A12-A14.
5. Atrial Fibrillation Investigators. Risk factors for stroke and efficacy of antithrombotic therapy in atrial fibrillation. Arch Intern Med 1994;154:1449-57.
6. McAlister FA, Straus SE, Guyatt GH, Haynes RB, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature. XX. Integrating research evidence with the care of the individual patient. JAMA 2000;283:2829-36.
7. User's Guides to the Medical Literature. A Manual for Evidence-Based Clinical Practice. Chicago: JAMA & Archives Journals AMA 2001.
8. Packer M, Coats AJS, Fowler MB, et al. For the Carvedilol Prospective Randomized Cumulative Survival Study Group. Effect of carvedilol on survival in severe chronic heart failure. N Engl J Med 2001;344:1651-8.
9. CIBIS-II Investigators and Committees. The Cardiac Insufficiency Bisoprolol Study II (CIBIS-II): a randomised trial. Lancet 1999;353:9-13.

10. Antiplatelet Trialists' Collaboration. Secondary prevention of vascular disease by prolonged antiplatelet treatment. BMJ 1988;296:320-31.
11. The Canadian Cooperative Study Group. A randomized trial of aspirin and sulfinpyrazone in threatened stroke. N Engl J Med 1978;299:53-9.
12. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. Lancet 2000;355:1064-9.
13. Lee DC, Johnson RA, Bingham JB, et al. Heart failure in outpatients: a randomized trial of digoxin *versus* placebo. N Engl J Med 1982;306:699-705.
14. Furberg CD, Byington RP. What do subgroup analyses reveal about differential response to beta-blocker therapy? The Beta-Blocker Heart Attack Trial experience. Circulation 1983;67:98-101.
15. The GUSTO investigators. An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. N Engl J Med 1993;329:673-82.
16. Furberg CD, Morgan TM. Lessons from overviews of cardiovascular trials. Stat Med 1987;6:295-306.
17. Oxman AD, Guyatt GH. A consumer's guide to subgroup analysis. Ann Int Med 1992;116:78-84.
18. Guyatt GH, Sullivan MJ, Fallen EL, et al. A controlled trial of digoxin in congestive heart failure. Am J Cardiol 1988;61:371-5.
19. Bulpitt CJ. Subgroup analysis. Lancet 1988;2:31-4.